

Presentation outline

Motivation

Poverty indicators

Data sources

Poznan agglomeration

The approach

Results

Measures of precision

Conclusions

Motivation

- ▶ demand for information concerning the quality of life, especially poverty, for low-level unplanned domains, such as the Poznan Agglomeration
- ▶ estimates at lower levels of the territorial division than regions (NUTS 1) or provinces (NUTS 2) have not been published so far
- ▶ such estimates can be calculated by using indirect estimation methods, which rely on out-of-sample data, and tend to improve estimation precision
- ▶ opportunity to draw on CSAE's experiences of poverty mapping:
 - for subregions - a project with the World Bank (2013)
 - for districts - an EU-funded POPT project (2015)

Poverty indicators

- ▶ **at-risk-of-poverty rate (head count ratio (HCR), poverty incidence** - the share of people with an equivalised disposable income (after social transfers) below the poverty threshold, (60 % of the national median EDI)
- ▶ **relative median at-risk-of-poverty gap** - the difference between the median EDI of people below the poverty threshold
- ▶ **Gini coefficient** - measure of income distribution inequality: ranges from 0 (maximal equality) to 1 (maximal inequality)
- ▶ **inequality of income distribution, income quintile share ratio (QSR), S80/S20 ratio** - the ratio of total income received by the richest top quintile of the population to that received by the bottom poorest quintile

Poverty indicators

Table 1: Poverty indicators for NTS 1 regions and Poland based on EU-SILC 2015

NTS1 region	Gini	HCR	QSR	Pov-Gap
central	34.0	16.5	5.7	-
southern	29.0	14.5	4.6	-
eastern	28.5	22.0	4.4	-
north-western	29.2	17.6	4.6	-
south-western	30.1	14.4	4.8	-
northern	29.7	20.5	4.7	-
Poland	30.6	17.6	4.9	22.3

Data sources used in the study (1)

The European Union Statistics on Income and Living Conditions (EU-SILC)

- ▶ two-stage stratified sampling with proportional allocation:
PSU - census output areas (stratified), SSU - households
- ▶ first sample in 2005 - 24,000 HHs, divided into 4 subsamples, each year one subsample replaced with a new one
- ▶ results published for domains defined by place of residence (urban/rural), locality size class, NTS1 regions, socio-demographic variables

Data sources used in the study (2)

The National Census of Population and Housing (2011)

- ▶ mixed mode census - registers + 20% sample
- ▶ **information available at NUTS2 level (provinces):**
 - economic activity by age group
 - unemployed people by duration of job search
 - economically inactive by cause of inactivity
 - employed people by employment status
- ▶ **information available at NUTS4 level (districts):**
 - economic activity by sex - level of education
 - marital status
 - sources of income
 - disability

Datasets and variables (1)

Table 2: Number of households in the two datasets
Wielkopolskie province and Poznań agglomeration

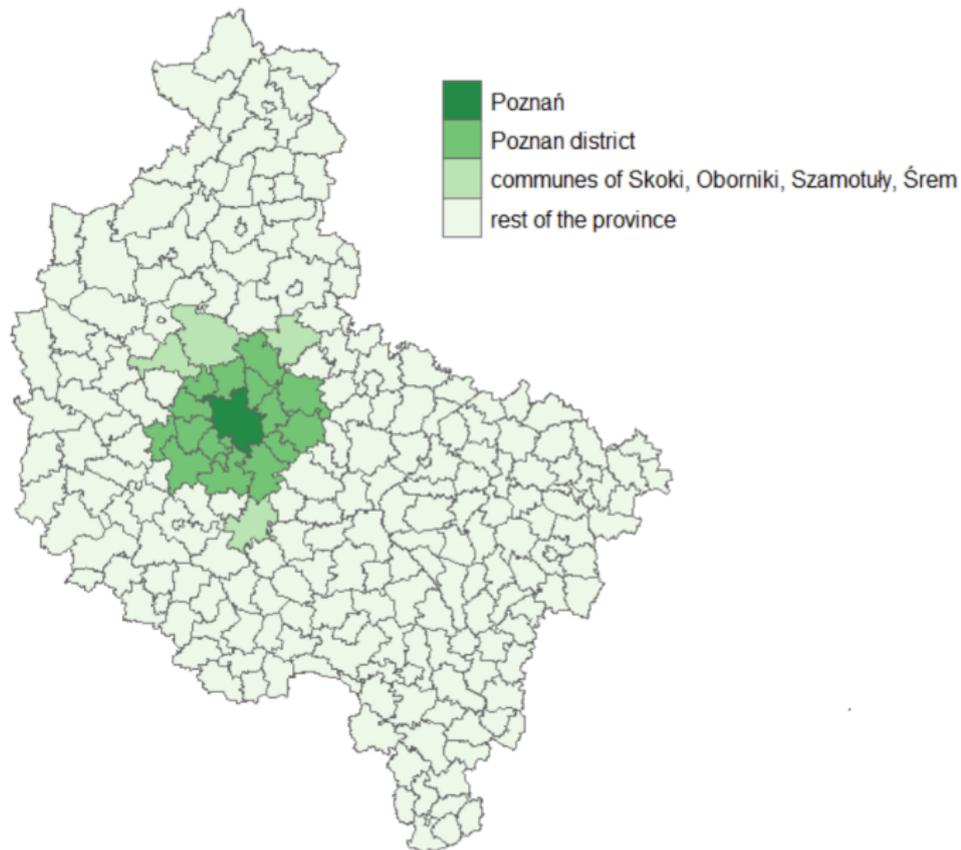
Dataset	province	agglomeration
EU-SILC 2015	821	173
Census 2011	228 632	40 840

Datasets and variables (2)

Household variables used in the model

- ▶ equivalised disposable income
- ▶ share of men
- ▶ share of people aged 65 or older
- ▶ share of unemployed persons
- ▶ share of disabled persons
- ▶ share of people with primary education
- ▶ share of people with higher education
- ▶ ratio of children (aged <15) to persons aged 16-65
- ▶ binary – HH in rural area or in a municipality < 20000
- ▶ binary – HH with 1 room
- ▶ binary – HH with 3 or more rooms

Poznan agglomeration



Poznan agglomeration

- ▶ created in 2007 as a way of fostering cooperation between the city of Poznan and 17 surrounding municipalities (the Poznan district)
- ▶ later on extended by including 4 municipalities from outside the Poznan district
- ▶ 11% of the province's area
- ▶ 30% of the total population (appr. 1 million)
- ▶ 40% of all companies registered in the province
- ▶ provide employment to 40% of all employees

The approach

1. estimate a **unit-level model** for the province of wielkopolskie (unit - household) with **random effect u_d at district level** (Poznan agglomeration is treated as one district)
2. using the model parameters, apply the EB method to predict values of **the dependent variable Y_{dj}** : equivalised disposable income in the household for out-of-sample units

The linear mixed model

$$Y_{dj} = x_{dj}^T \beta + u_d + e_{dj}, \quad j = 1, \dots, N_d, \quad d = 1, \dots, D,$$

where:

Y_{dj} - Box-Cox transformed income of j -th HH in district d

x_{dj}^T - vector of independent variables for j -th HH in district d

β - vector of regression parameters

u_d - random area effect with $u_d \stackrel{iid}{\sim} N(0, \sigma_u^2)$

e_{dj} - residual errors $e_{dj} \stackrel{iid}{\sim} N(0, \sigma_e^2)$

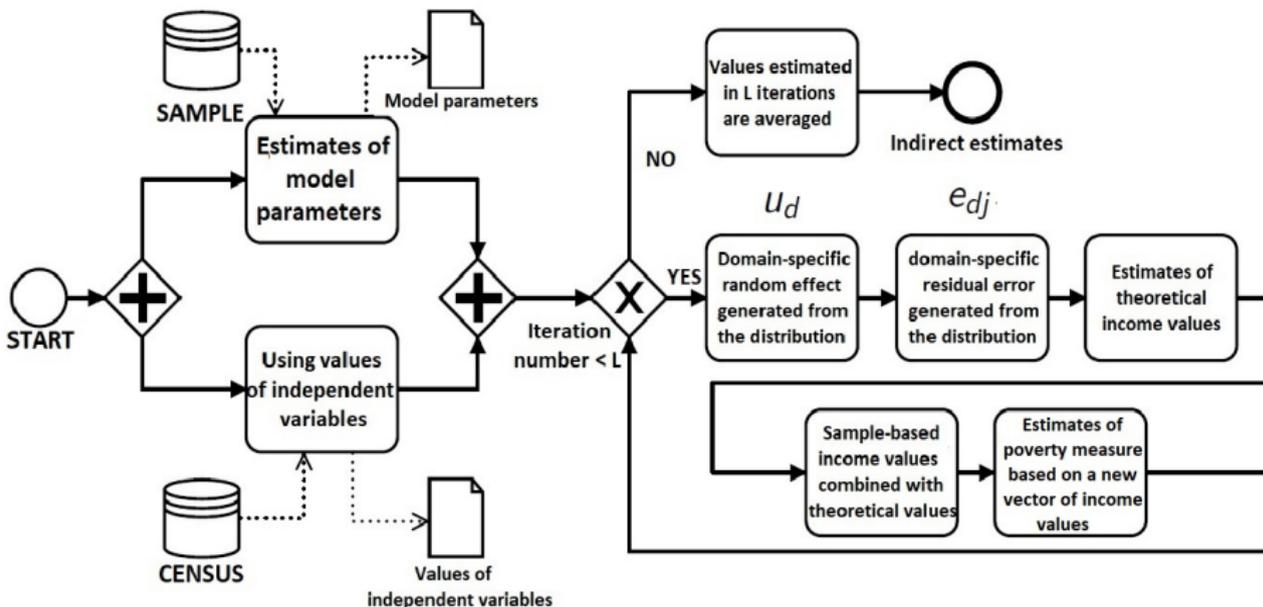
Calculations made using the `lmer` function from the `lme4` package in R.

Empirical Best Prediction (Molina-Rao 2010)

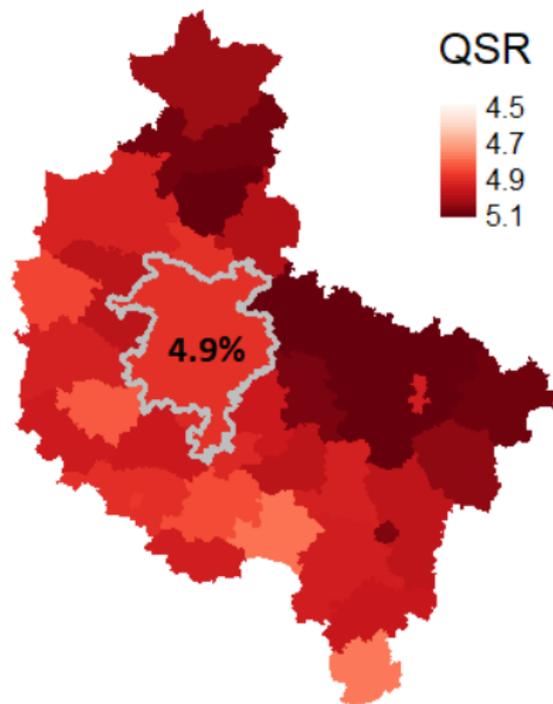
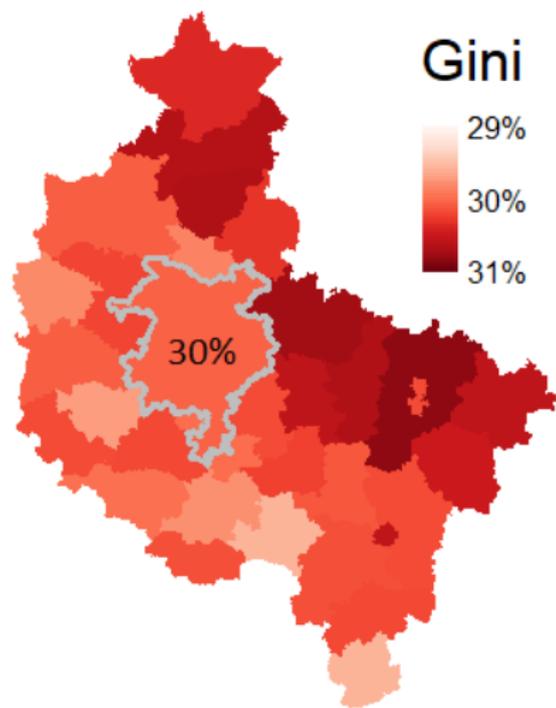
1. fit a linear mixed model describing the household income based on (transformed) data from a representative survey
- y_s
2. draw L out-of-sample vectors y_r but with the unknown parameters replaced by the estimators obtained in (1) - model parameters and random effects variances
3. With the L generated vectors of theoretical values and using the sample data y_s , compute EBPs of the poverty measures for unit-level auxiliary data from the census using the Monte Carlo approximation (minimizing MSE under the model).
4. estimate MSE using parametric bootstrap

Calculations made using the [ebp](#) function from the [emdi](#) package in R (key parameters: $L = 500$, $B = 200$)

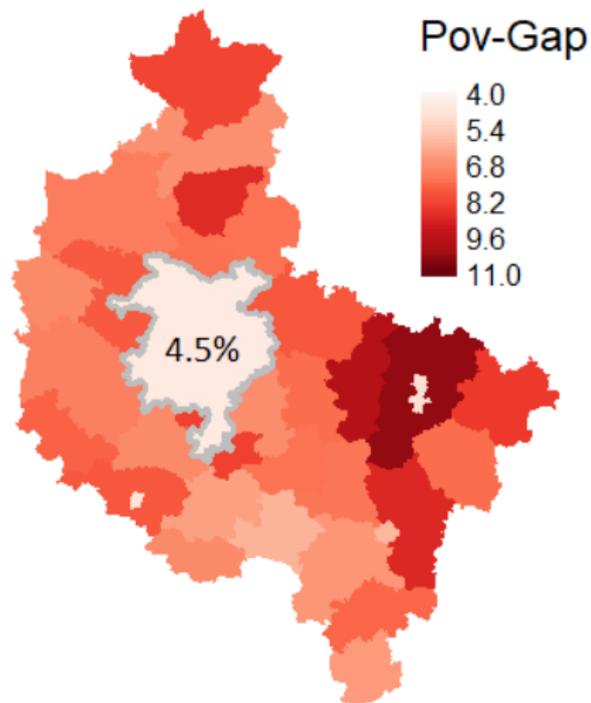
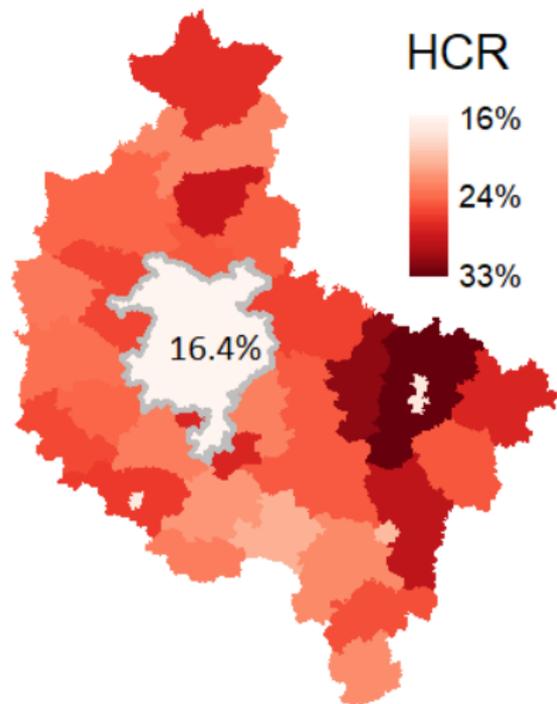
The estimation process



Results: Gini coefficient & Quintile share ratio



Results: HCR (poverty rate) & Poverty gap



Measures of precision

Parametric bootstrap for MSE estimation - approximated through a Monte Carlo procedure repeated B times and then taking the mean over the B replicates.

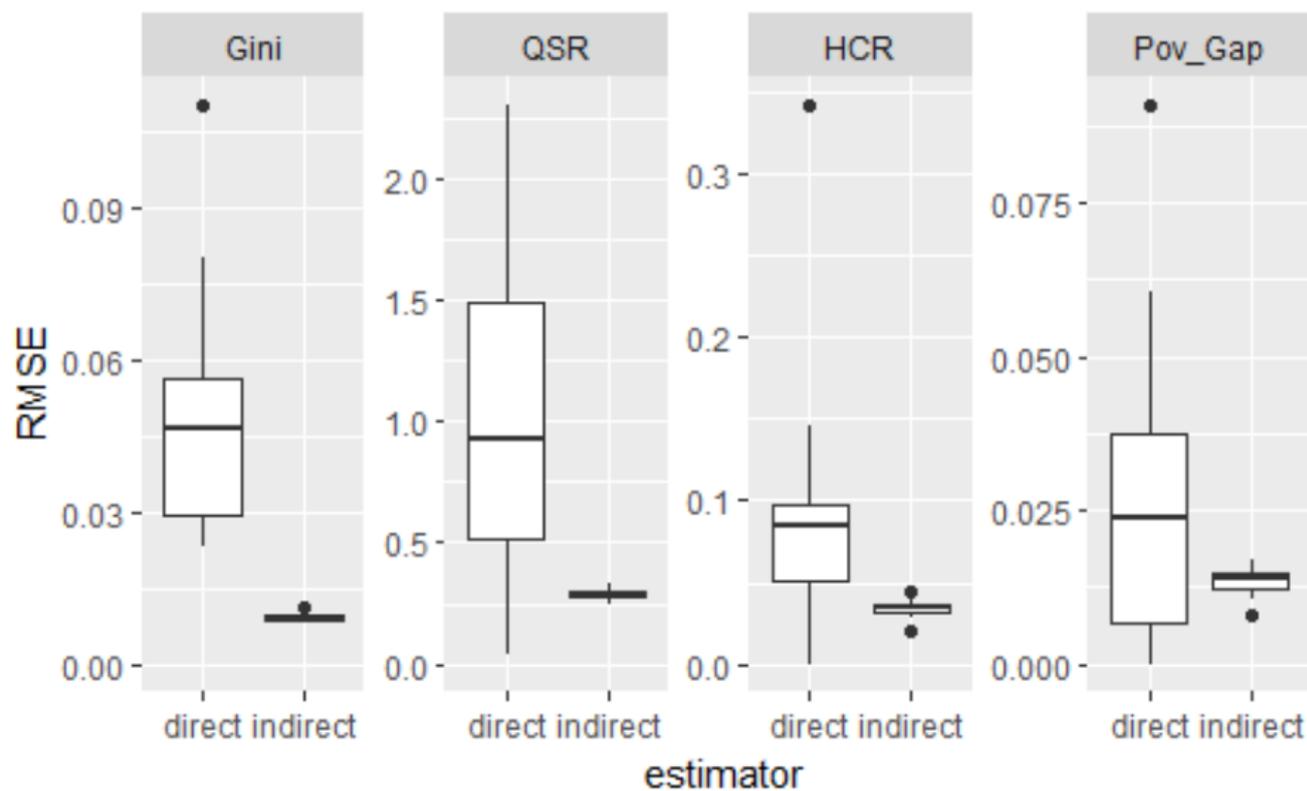
$$mse(\hat{Y}_d^{EB}) = \frac{1}{B} \sum_{b=1}^B (\hat{Y}_d^{EB*(b)} - Y_d^{(b)*})^2$$

Relative root mean square error (CV) (%):

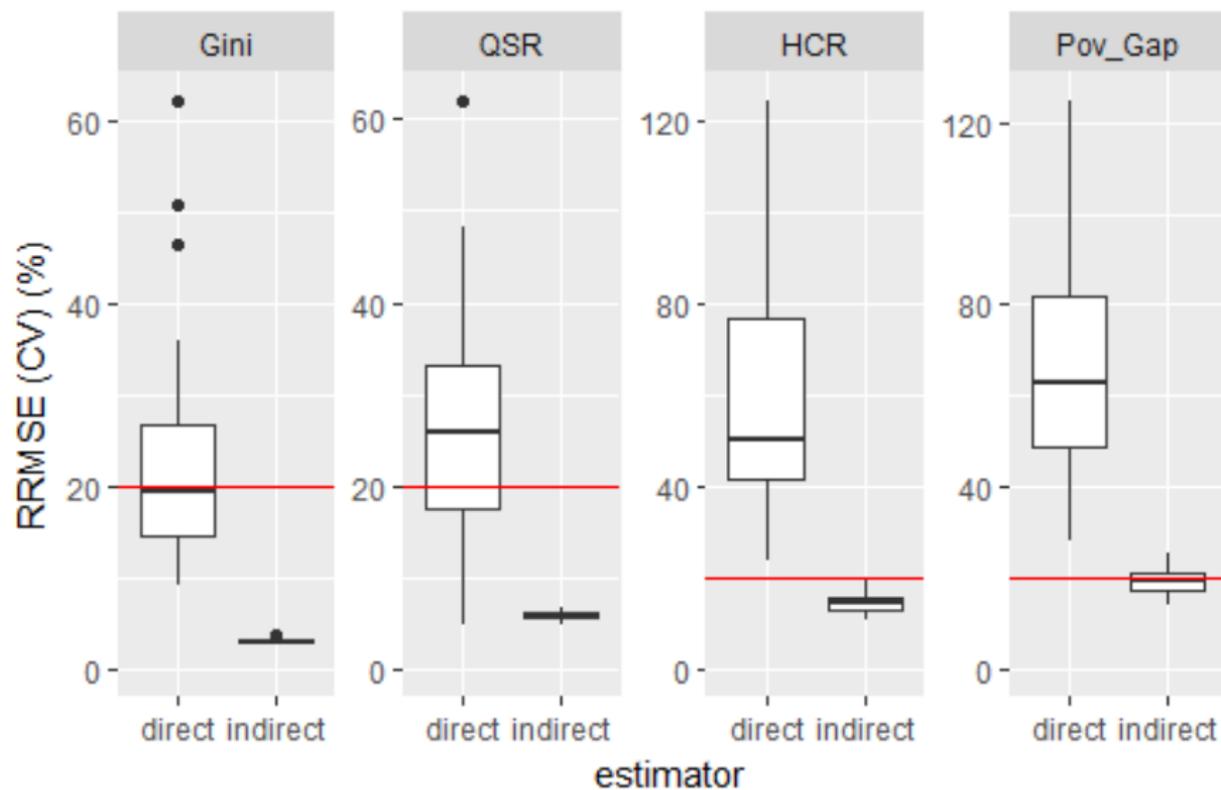
$$RRMSE(\hat{y}) = \frac{RMSE(\hat{y})}{\hat{y}} * 100$$

\hat{y} - estimate of a poverty measure

Root mean square error



Relative root mean square error (CV) (%)



Conclusions

- ▶ methods of indirect estimation make it possible to estimate poverty for unplanned domains (when sample size is small or even zero)
- ▶ further work is required to select the most optimal explanatory variables for the model (especially variables from administrative registers, including the tax register)
- ▶ estimation should be conducted for subsequent years to verify the model's stability over time

References (selected)

- ▶ GUS, (2017), *Income and living conditions of the population of Poland (report from the EU-SILC survey if 2015)*, Warszawa
- ▶ Molina, Isabel and Rao, J. N. K, (2010), Small Area Estimation of poverty indicators, *Canadian Journal of Statistics*, 38(3):369–385.
- ▶ Wawrowski, Ł. (2017), Indirect estimation of poverty at the regional and local level in Poland, (doctoral dissertation)

Thank you for your attention.



**Urząd Statystyczny
w Poznaniu**